Preserving Early UNIX* Systems

Warren Toomey
Email: wkt@cs.adfa.edu.au

Printed January 21, 1999

Abstract

The UNIX operating system has been, and continues to be, a very influential operating system. This paper reviews the efforts being undertaken to preserve old versions of UNIX (both source and binaries), to gather the information required to maintain systems running these versions of UNIX, and the progress that has been achieved in providing legal access to old UNIX source by individuals.

WKT – We want to describe what we are doing with respect to UNIX. This includes everything: finding tapes, building emulators, fixing old versions, helping each other out, scanning paper documents etc. We should not just mention the PDP-11 work.

We should also put footnotes in (for the first occurrence) that say the terms PDP and VAX are trademarks of DEC? Compaq?!

1 Introduction

The UNIX operating system, conceived in 1969 and developed since then, has had a remarkable effect on both computing research and the general computing industry. Innumerable versions and derivatives of UNIX have been produced, and many of its features have been taken up in other systems.

Figure 1: Dennis Ritchie & Ken Thompson at the PDP-11/20 console

The history of UNIX, and its influence on computing, has been well documented [1] [2] [3] [4] [5]. But although the *story* of UNIX has been captured to some extent, there has been little effort to try and preserve the *artifacts* in the history of UNIX: its source code, documentation, system binaries, and the reports and papers written about it.

WKT – Is there a need at this point for a brief recapitulation of the history of UNIX? Some readers may not have any knowledge of the history.

2 The UNIX Heritage Society

The UNIX Heritage Society is a very loose affiliation of people involved in the preservation, maintenance and development of UNIX. Originally created in 1995 to preserve the PDP-11 versions of UNIX, it has since been extended to cover all old versions of UNIX (both research and commercial flavours).

---

*UNIX is a registered trademark of The Open Group

The Society has many aims, and include (but are not limited to):

- The preservation and maintenance of historical and orphaned UNIX systems; and
- The further development of existing UNIX systems.

Membership of the Society is not formal: any person who belongs to the Society's electronic mailing list, or who is associated with the Society's aims, are considered to be a member. The mailing list has 100 subscribers and continues to grow. It is mainly used to announce additions to the archive of software, to ask about the history of UNIX, and to ask about the maintenance of old hardware and the installation of UNIX onto this hardware. In this way, the mailing list members help each other out with information and advice which would be next to impossible to find.

3 The PUPS Archive

A central part of the Society is the PUPS[1] Archive: this is a repository of source, applications, information and other miscellaneous files which are related to UNIX and its history. The archive is kept as a set of files on disk: where information has been transcribed from tapes, we attempt to document and maintain the record structure of the tapes, and transcribe any printed labels from the tapes to text files.

The archive started off with a small number of old backup tapes from the late 1970s, none of which had complete UNIX distributions. As the mailing list grew, and more people became aware of our preservation efforts, many other tapes and files were added to the archive.

3.1 Recovering UNIX-Related Data

There are several obstacles to the recovery and preservation of digital data relevant to the history of UNIX: there has been no systematic preservation of data, data has been stored on media which is now difficult to read, and the format and structure of such data is often no longer documented.

The recovery of machine-readable data from the very early days of UNIX is hindered by the fact that it was a research project undergoing continuous development at Bell Laboratories. Although backups of the development were made periodically, these were not seen as important to keep for posterity. For the period 1969 to 1973, there is no extant machine-readable UNIX artifacts, except one file containing the processing routine for an assembly-language version of UNIX, dated 1972 [6]. *WKT – I need to verify with Dennis that this, indeed, is the oldest machine-readable file (because he also has tapes labelled 's1', 's2' and 'unix' which might be older. I also need to get a more exact date for this file,* `tty.s`. There are, however, earlier paper listings of the assembly-language version of UNIX for the PDP-11, and somewhat more fragmentary listings of the PDP-7 version.

The earliest significant set of UNIX files still in existence contain the source code to one of the first UNIX kernels written in C, and are dated January 23rd, 1973. At this point in the development of UNIX, the system was being developed concurrently in both C and PDP-11 assembly code; it was not until September 1973 that the C version of the UNIX kernel replaced the assembly-language version [7]. Despite the lack of a complete UNIX system from this date, members of the UNIX Heritage Society have been able to compile this kernel, and boot it on a PDP-11/45 platform.

The loss of machine-readable data from the development of UNIX continues up until March 1975. At this date, there exists a nearly-complete copy of the 5th Edition of UNIX, with only the documentation missing. This system is in a state that allows it to be installed and booted on a PDP-11/45 platform.

After the publication of "The UNIX Time-Sharing System" [1] in 1974, UNIX was distributed to institutions outside of Bell Laboratories. The chance of recovering machine-readable data from this point onwards is therefore much greater.

3.2 Reading Old Media

WKT – A section on difficulties in reading old media. Perhaps some input from Tim Shoppa here. During the early development of UNIX, filesystems were stored on, and backups made to, various media: 800bpi and 1600bpi 1/4" magnetic tape, DECtape, Baudot-encoded paper tape, proprietary removable or fixed hard disk storage units.

3.3 Interpreting Data Formats

Even if the contents of a disk or tape can be read, there remains the issue of interpreting the data that it contains. Fortunately, the number of filesystem and backup formats used by the various versions of UNIX is relatively small, and nearly all are documented.

---

[1] The archive retains the acronym of the Society's original name: the PDP UNIX Preservation Society.

Early UNIX systems on the PDP-11 family often had removable disk packs; backups were often performed by copying one filesystem to an empty disk pack. The format of the filesystem changed several times from its creation through to the 7th Edition in 1979, and then again as it was modified by other institutions (for example, the Berkeley Fast File System) [8]. The format for each filesystem is well documented, and the Society has a number of programs to extract files from filesystem images; the UNIX attributes for each file (owner, group, permissions, timestamps) are also preserved.

Recovering data from UNIX tape archives is much more interesting, due to the wider variety of formats and their interpretation. The archive format dating from 1st Edition UNIX (1971) is *tap*. With *tap*, the times associated with the files were recorded in 32-bit fields holding sixtieths of a second from an origin that changed. As this field overflows after 2.25 years, the origin changed at least once (from 1971 to 1972), in the period that *tap* was in use [6]. This makes the interpretation of the dates in *tap* archives troublesome.

By November 1973, *tap* had been replaced by the *tp* format, where dates are stored in seconds [6]. This format remained in used until the 7th Edition (1979), when it was replaced by the *tar* format, which is still in used today, albeit modified somewhat. The *cpio* backup format was introduced in the Programmer's Workbench UNIX in 1977, and is also still in use today. The Society has programs to read all of these tape formats.

3.4 The Contents of the PUPS Archive

WKT – I would like to change the following, probably to make it more like prose rather than an itemised list.

At present, the PUPS Archive contains:

Bell Labs UNIX: 5th Edition, 6th Edition, 7th Edition, V7 patches.

USDL/USL: PWB/UNIX, Mini-UNIX, 32V, System III for the PDP-11, System V for the PDP-11.

CSRG: 1BSD, all the 2BSDs, 3BSD, all the 4BSDs including 4.4, Net/1 and Net/2.

DEC: V7M, Ultrix-11 3.0 and 3.1.

Homegrown UNIX: AUSAM, other modified versions.

Software: Several old USENIX tapes supplied by Henry Spencer; the Software Tools from Debbie Scherrer.

Bug Fixes: An archive of the Usenet newsgroup `net.v7bugs` from Henry Spencer; other archives of bug fixes for 2BSD and 4BSD.

Tools: Several tools for dealing with tape archive formats and filesystem formats from early versions of UNIX.

Emulators: Several emulators for early UNIX hardware platforms.

Unfortunately for future historians, nearly all of the details of incremental changes between UNIX versions has been lost. The exception is 4BSD. The crew at the CSRG at Berkeley began using SCCS to control the revisions on their source code from 1980, and so (with some exceptions) the development of 4BSD from this date onwards can be traced.

4 Preserving the Hardware Environment

WKT – anybody care to give a few paragraphs about the care and feeding of old computers, in particular the main UNIX platforms: PDP-11 and VAX?

The hardware emulators in the archive are also very important, for although some people own old computers, most people in the Society do not. The emulators allow them to experience the environments provided by early UNIX despite the lack of suitable hardware. There are several PDP-11 emulators currently in the archive. As yet, there is no Vax emulator, but at least one PUPS member is planning on constructing such an emulator.

5 Improving and Extending Older UNIX Systems

WKT – We should try to get Steven Schultz to give some words here about his long-standing work on 2.11BSD. Will replace the next short paragraph.

Steven Schultz took over the maintenance of 2BSD (for the PDP-11) in the 80s, and is still developing it. The current version, 2.11BSD, has most of the functionality of 4.4BSD, except for a much smaller process address space. If you use any current BSD system, then you would find yourself at home on 2.11BSD.

Another effort is the Quasijarus project, led by Michael Sokolov. The aim here is to extend the support of 4.3BSD (and 4.3BSD-Tahoe) to allow it to be run on nearly all of the VAX systems in the VAX minicomputer family.

6 Providing Access to UNIX Source

In order to access the source code of a UNIX system, one still requires a UNIX source license. This includes all versions of UNIX from Bell Labs, AT&T, USL, SCO, and all flavours of Unix that are tainted with UNIX source code, e.g Net/2 and 4.4BSD. The source code to Unix systems that are untainted, e.g Linux and the free BSDs, do not require a UNIX source license.

When the UNIX Heritage Society was first started, it was impossible to buy a source license for the historical versions of UNIX that we were interested in preserving. The only way of legally possessing UNIX source was to be an employee of an institution covered by an existing source license, or to buy a new UNIX source license costing tens of thousands of dollars.

Just before the 25th anniversary of UNIX in 1995, John Lions and Peter Salus began to lobby Novell (who owned the UNIX source at the time) to allow Lions' commentary on the 6th Edition [9] to be properly published. It was not until the UNIX source was sold to SCO that Salus (with Dennis Ritchie's help) persuaded SCO to release the Lions' commentary. This was finally published in July 1996 by Peer to Peer Communications and includes most of the kernel source to 6th Edition. The copyright does not allow the source code in the book to be converted to machine-readable format.

In early 1996, Steven Schultz and I began to lobby SCO for a solution. Initially, we hoped to get SCO to make early UNIX source freely available, but this was not to be. SCO believed that this would dilute the copyright on, and the trade secrets within, the current versions of UNIX produced by SCO. Therefore, a new type of source license was the only option.

We were lucky to contact Dion Johnson at SCO, who appreciated our wish for easy access to the old UNIX source, and our desire to abide by the conditions of a negotiated source license. In mid-1996 we conducted a Usenet survey to find out what conditions and costs in a license would be acceptable to most people. This was followed up by an on-line petition to SCO in July 1997, to make cheap licenses available. Over 300 people signed the petition.

Within SCO, Dion Johnson did the hard work of keeping the pressure on the legal eagles to create a source license for early UNIXes. Finally, in March 1998, SCO released their "Ancient UNIX" license. This costs US$100 and provides an individual, non-commercial-use, source license for:

- 1st Edition up to 7th Edition UNIX, inclusive;
- 32V; and
- Other PDP-11 UNIXes owned by SCO, such as PWB/UNIX and Mini-UNIX.

Having 32V included on the license was critical, because a source license for 32V is required if you wish to obtain any of the tainted Berkeley Software Releases from CSRG.

Since the release of this license by SCO, over 120 people have bought licenses. SCO does not distribute any software with the licenses; instead, the Society endeavors to distribute the software covered by the SCO license to licensees if they want it.

7 Distributing the PUPS Archive

The UNIX Heritage Society provides access to the PUPS Archive to UNIX source license holders in two main ways. UNIX source license holders can request access to the password-protected on-line version of the PUPS Archive. The on-line PUPS Archive is the canonical, up-to-date version of the archive. Currently, 65 people have on-line access to the archive.

The Society also distributes copies of the PUPS Archive on CD-ROM, magnetic tape or other media. We have eighteen people around the world who can duplicate CDs or tapes, and post them to licensees. This distribution is done on a cost-recovery basis, and so the extra cost on top of the SCO license is minimal.

Related to the PUPS Archive distribution is the four CD set that Kirk McKusick sells. This contains all of the BSD software releases from the CSRG from 1BSD up to 4.4BSD-Lite2. Although all of this material as part of the PUPS Archive, it is not included on-line or distributed by CD, so as to not compete with Kirk's CD set.

8 Old Artwork

The UNIX Heritage Society is interested in collecting anything that is related to old UNIXes, and this includes UNIX-related artwork.

One of the earliest pieces of UNIX art was done in 1976 by artist Phil Foglio. It shows several daemons holding forks on a PDP-11 festooned with pipes and a bucket labelled *NULL*. Unfortunately, this artwork has since been mislaid. Phil Foglio originally drew the artwork for Mike O'Brien, who then lent it to Armando Stettner at DEC for use in an advert in the 1980s. Neither Armando nor Jon Hall at DEC have been able to track what happened to it after that.

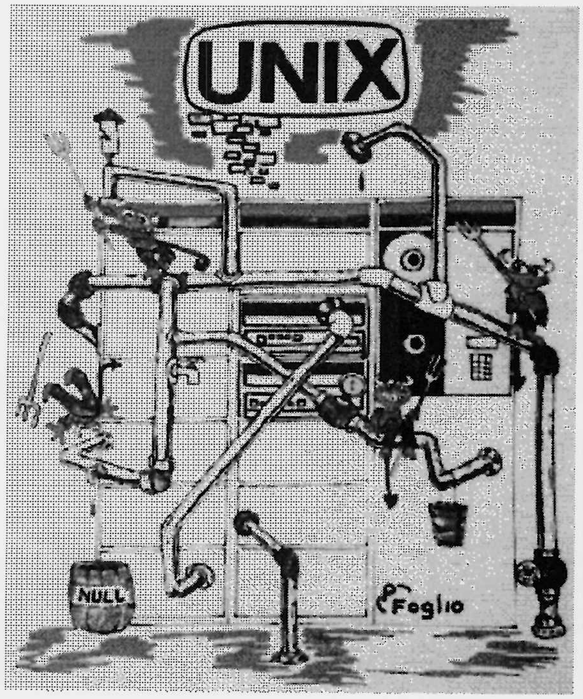

Figure 2: Phil Foglio's 1976 artwork

Many other photos, drawings, diagrams and works of UNIX art are probably still lying in people's drawers or shoeboxes. PUPS would certainly be glad to add digital copies of these to the PUPS Archive, with due credit given to the owners of the original images.

9 Old Documentation

Although most of the UNIX source code from before 5th Edition has disappeared, it is fortunate that paper copies of all the research edition manuals still exist. Dennis Ritchie has scanned in the paper manuals from 1st Edition UNIX and put them on his web page. Two other Society members, Norman Wilson and Robert D. Keys, have been OCR'ing the manuals from 1st Edition up to 5th Edition, so that they can be given to Dennis and added to the PUPS Archive.

10 The UNIX Family Tree

One other Society project is the construction of a UNIX family tree. Access to the PUPS Archive here is vital, as it provides accurate information about the release dates and genealogy for each of the flavours of UNIX that have been released. What is lacking in the current tree is information about commercial versions of UNIX such as AIX, Xenix, Ultrix, Solaris and the like.

11 Conclusion

To many people, preserving early versions of UNIX might seem to be a strange pastime. However, those who collect or restore old books, cars, antiques or other collectibles might understand what motivates the effort to save UNIX from `/dev/null`.

Since the creation of the UNIX Heritage Society in 1995, we have acquired an large archive of material from the early days of UNIX, including many early versions. We now have a loose network of over 100 people who can help each other out with questions about old UNIX systems and old hardware. One of our most important achievements is to have successfully lobbied SCO to release cheap source licenses for many of these early versions of UNIX.

Preservation of the UNIX legacy will be an ongoing effort. We have made a good start at capturing what exists today. As we move into the new century, the emphasis of PUPS will shift move to preserving systems from the 1980s and 1990s.

The UNIX Heritage Society will always be interested in the history and artifacts of UNIX. If you have anything which you think might belong in our archive, or you know someone who might have something to add, then please contact us so that we can save it before it is irrevocably lost.

12 Acknowledgments

WKT – Anybody we should thank, or who has made a significant contribution to our efforts, *should be mentioned here.*

Fortunately, there are many pack rats out there in the UNIX community. Thanks must go to Dennis Ritchie, Kirk McKusick, Henry Spencer, Keith Bostic, Steven Schultz, Ken Wellsch and others who have made significant contributions to the archive.

References

[1] D. M. Ritchie and K. Thompson. The unix time-sharing system. *Comm. Assoc. Comp. Mach.*, 17(7):365–375, July 1974.

[2] D. M. Ritchie. The evolution of the unix time-sharing system. In *Proc. of Symp. on Language Design & Programming Methodology*, Sydney, September 1979. Also in *BLTJ*, 63 (8, Part 2), pp. 1897-1910, October, 1984.

[3] P. Salus. *A Quarter Century of UNIX*. Addison-Wesley, 1994. ISBN 0-201-54777-5.

[4] D. Libes and S. Ressler. *Life with UNIX*. Prentice-Hall, 1991. ISBN 0-13-536657-7.

[5] M. Hauben and R. Hauben. *Netizens: On the History and Impact of Usenet and the Internet*, chapter 9. IEEE Computer Society, 1997. 0-8186-7706-6.

[6] D. M. Ritchie. Private email to W. Toomey, dated Sep 23 1997.

[7] D. M. Ritchie. Private email to W. Toomey, dated Jan 19 1999.

[8] M. McKusick, W. Joy, S. Leffler, and R. Fabry. A fast file system for unix. *ACM Transactions on Computer Systems*, 2(3):181–197, August 1984.

[9] J. Lions. *Lions' Commentary on UNIX 6th Edition, with Source Code*. Peer-to-Peer Communications, 1996. ISBN 1-57398-013-7.